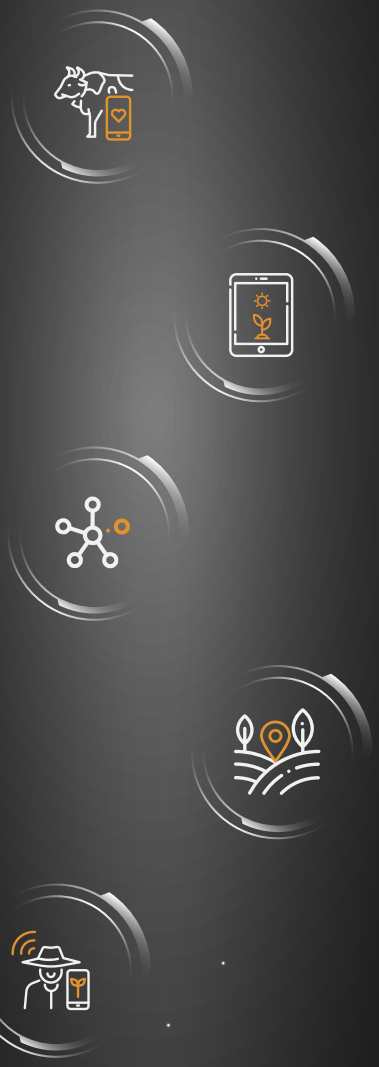
CGIAR

Platform for Big Data in Agriculture

# CGIAR PLATFORM FOR BIG DATA IN AGRICULTURE **ANNUAL REPORT** 2020



Alliance

Bioversity International

CIAT
International Center for Tropical Agriculture
Since 1967 Science to cultivate change

IFPRI
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE


**Name of the Platform:**
BigData

**Name of Lead Center:**
CIAT (Alliance)

**Module lead institutions (CGIAR Centers or lead partners)**

**Module 1:** Organize
**Module 2:** Convene
**Module 3:** Inspire

**Other participating Centers:**
AfricaRice, Bioversity (Alliance), CIAT (Alliance), CIFOR, CIMMYT, CIP, ICARDA, ICRAF, ICRISAT, IFPRI, IITA, ILRI, IRRI, IWMI, WorldFish

# TABLE OF CONTENTS

# EXECUTIVE SUMMARY

The Platform completed its fourth year as a CGIAR global digital practice, demonstrating that open data standards, sharing, and analytic infrastructure; partnerships and communities of practice; and applied innovation strategy reinforce each other to accelerate inclusive, impactful digitization in agricultural development. In 2020, BIG DATA brokered **72 external partnerships** to drive digital innovation in research and unlock data sharing across public, private, and non-profit organizations; and delivered **22 digital innovations** and policies to build new digital impact pathways for CGIAR research.

Through its programming, partnerships, and strategic research, BIG DATA demonstrated that mission-driven "earth shots" in food security are achievable through digitally-enabled innovation and collective action, and that a unified CGIAR can play a key role in achieving them.

## ORGANIZE

Under its Organize Module, the Platform continued to build the global knowledge base for agricultural research through its Global Agricultural Research Data Innovation and Acceleration Network (GARDIAN). The information assets shared by all CGIAR centers and a growing number of strategic partners via GARDIAN reached a total of **191,570 publications** and 38,735 datasets. GARDIAN's analytic environment, CG Labs, helped researchers generate valuable new analyses through collaboration, analysis, and visualization, aided by pre-packaged datasets, scripts, and tooling. Researchers across CGIAR used CG Labs to create analytic products on: (1) Climate risks to agriculture in 16 countries; and (2) Analysis of the return on investment of fertilization in Sub-Saharan Africa, using more than 200 CGIAR datasets. The team revised the CGIAR Open Access/Open Data policy to reinforce the Findable, Accessible, Interoperable, and Reusable (FAIR) principles for use in CGIAR's and its partners' data and publications, building momentum to ensure agricultural research for development data is re-usable.

## CONVENE

The Convene Module developed CGIAR thought leadership through seven technical Communities of Practice (CoPs), held a virtual convention that reached millions of stakeholders and enabled the formation and fostering of new digital alliances. The CoPs grew to more than 5,000 members, becoming a key means for CGIAR to identify sector bottleneck issues and develop digital research innovations. It released important community-driven products including a tool to assist crop breeders in trait selection, a pipeline for delivering daily weather data to locations in the International Wheat Improvement Network, a household survey ontology, and a system for calculating greenhouse gas (GHG) emissions from multiple crop events. These research products were disseminated through 45 CoP-led webinars and workshops to more than 5,600 live participants. BIG DATA published guidance for the sector on building enabling data ecosystems for startups, technology standards for protecting agricultural data in information systems, and the potential risks of widespread use of Artificial Intelligence (AI) in agricultural value chains.

## INSPIRE

The Inspire Module awarded seven start-up grants in the CGIAR digital innovation program—the Inspire Challenge—and grew its portfolio of active projects to 21. The team made three "rapid response grants" to previous awardees to leverage their projects for COVID-19 response, recovery, and resilience. The Challenge attracted US$500,000 in external funding in 2020.

01

# KEY **RESULTS**

## 1.1 HIGHLIGHT PLATFORM ACHIEVEMENTS

The Platform completed its fourth year of building CGIAR's cross-cutting digital capabilities, partnerships, and innovation strategy in digital agriculture. The team formed 72 external partnerships to drive digital innovation in research and data sharing; and delivered 22 innovations and policies to build new digital impact pathways for CGIAR research.

Through its program, partnerships, and wide-ranging strategic research in 2020, BIG DATA demonstrated that mission-driven "earth shots" in food security are achievable through digitally-enabled innovation and collective action, and that a unified CGIAR can play a key role in achieving them.

### ORGANIZE

Under the Organize Module, CGIAR and a growing number of strategic partners shared a cumulative 191,570 publications and 38,735 datasets via GARDIAN, CGIAR's flagship data harvester. The Platform led the revision of the organization's Open Access/Open Data Policy to elevate FAIR (findable, accessible, interoperable, and reusable) principles for CGIAR and partners' data and publications, helping build the momentum for the practice of creating reusable data in agricultural research. Research teams used GARDIAN's analytic workbench, known as CG Labs, to analyze large, complex datasets for two banner products in 2020. All scripts that were developed by BIG DATA to clean, process, and standardize datasets are open and available for reuse.

### CONVENE

Under the Convene Module, CGIAR built its thought leadership through seven technical Communities of Practice (CoPs) sponsoring a virtual convention that reached millions of stakeholders, and forming new digital alliances. The CoPs grew to more than 5,000 members, becoming a key means for CGIAR to stay abreast of and develop digital research innovations. The CoPs released important community-driven products including a tool to assist crop breeders in trait selection, a pipeline for delivering daily weather data to locations in the International Wheat Improvement Network, an ontology for household survey data, and a system for calculating greenhouse gas emissions from multiple crop events. These outputs were disseminated through 43 webinars and workshops. BIG DATA issued guidance for the sector guidance on building data ecosystems for startups, technology standards for protecting agricultural data, and the risks of using AI in agricultural value chains.

### INSPIRE

The Platform awarded seven start-up grants through its digital innovation process, the Inspire Challenge, growing the portfolio to 21 active projects and made three "rapid response grants" for COVID-19 response, recovery, and resilience. The Module attracted US$500,000 in new external funding in 2020.

# BIG DATA AT A GLANCE

## INNOVATIONS

US$ **800K** IN GRANTS

**7** INSPIRE CHALLENGE **PROJECTS**

**3** RAPID RESPONSE **GRANTS**

**21** ACTIVE INSPIRE **PROJECTS**

## COVID-19

US$ **100K** AWARDED IN RAPID RESPONSE GRANTS

**3** **PROJECTS MITIGATING COVID 19 IMPACTS**

**DIGITAL DISCUSSION SERIES ON COVID-19** IMPACTS - CHALLENGES & OPPORTUNITIES

**7** EPISODES

**22** SPEAKERS

**4,300** VIEWERS

## PARTNERSHIPS & COLLECTIVE ACTION

MORE THAN **70** **PARTNERSHIPS**

**BIG DATA** COMMUNITIES OF PRACTICE GREW **67%** TO **5000** MEMBERS

**INSPIRE CHALLENGE & RAPID RESPONSE GRANT PARTNERSHIPS**

**9** **CGIAR** CENTERS AND **RESEARCH** PROGRAMS

**22** EXTERNAL PARTNERS

## COMMUNICATIONS / OUTREACH

BIG DATA WEBSITE

**200,000** VISITS
44% INCREASE IN TRAFFIC

**65,000** USERS
51% INCREASE

**SOCIAL MEDIA** INCREASED AUDIENCE **97%** TO **20.6K** **FOLLOWERS**

**BIG DATA COPS HELD 45 CAPACITY BUILDING WEBINARS**

**5,613** **LIVE VIEWS ON YOUTUBE**

**11,763** REPRODUCTIONS **ON YOUTUBE**

## 1.2 PLATFORM PROGRESS TOWARDS OUTPUTS AND OUTCOMES

### 1.2.1 OVERALL PLATFORM PROGRESS

The Platform continued to build CGIAR's digital capabilities, innovations, and partnerships for impact while navigating a fluid organizational environment as well as fluid conditions in CGIAR and in global food systems. In some ways, the COVID-19 crisis has buoyed the Platform's efforts. The crisis precipitated rapid digitization in the agri-food sector and highlighted the importance of digital tools for agile adaptation in food systems. It also has thrown disparities in the access to and use of these digital tools into sharp relief. Wide-ranging interviews conducted by the Platform in 2020 showed dramatic, rapid shifts in digital culture and practice in public, private, and non-profit food security organizations including the rapid digitization of organizational and research processes, a spike in demand for good quality data, increased use of computational research methods, and an abrupt move to use remote agricultural advisory services.

The COVID-19 crisis increased the global focus on data sharing, digitally-enabled adaptation, and computational research, and this accelerated the formation of two high-profile new alliances. CGIAR scientists joined X (formerly Google X) Digital Green, Plantix, and One Acre Fund in mobilizing crop data and developing and running a national-scale, dynamic, remote-sensing-based model for detecting maize crop emergence and the total in-season planted area in Kenya. This was implemented with unprecedented speed and at a national scale, demonstrating new methods for food security monitoring that will be needed for managing future crises. COVID-19 lockdowns provided a unique opportunity to detect gas emissions patterns, and CGIAR researchers and Hewlett Packard Enterprise developed a remote sensing model for detecting levels of nitrogen dioxide (NO2) emissions from space, creating valuable new data assets relevant to modeling the interactions between economic activity and climate.

Platform Modules continued to reinforce each other and the mission to digitally-enable CGIAR and its partners in 2020. Those researchers and partners engaged via the Convene Module, for example, began using the data discovery and analytic infrastructure developed under the Organize Module as the Platform mobilizes the data that is now more available and builds momentum for FAIR data standards use by CGIAR and the sector. The Inspire Module continues to spark new digital innovation partnerships and overall engagement with the Platform, including through its growing and vibrant technical CoPs (under Convene).

The BIG DATA team formed 72 external partnerships to drive digital innovation in research and unlock agricultural data sharing among public, private, and non-profit partners; grew its technical CoPs to more than 5,000 members; delivered 22 digital innovations and policies to build new digital impact pathways for CGIAR research; a doubled its social media audience to 20,000 across Twitter, LinkedIn, YouTube and Instagram; and increased its website traffic by more than 50% to 65,000 users.

In 2020, BIG DATA hosted the first One CGIAR Convention, during which all 13 CGIAR centers participated in some capacity. COVID-19 presented the Platform with a unique opportunity to "walk the talk" on agile, adaptive, digitally-enabled collective action, with BIG DATA transitioning its fourth annual convention to one that ran 100% virtually online. The event theme, Digital Dynamism for Adaptive Food Systems, examined food system resilience and highlighted how digital tools and technologies can help food system actors sense, respond, and (re)build better systems in times of global food security crises. With more than 2,500 registrants and 1,300 active participants, it was the most highly attended, inclusive, and accessible event created by the Platform to date.

Against this backdrop, the Platform led research for development of a One CGIAR digital strategy. The team consulted a wide array of stakeholders. In total, it conducted 165 surveys and 80 semi-structured expert interviews, and led 10 internal workshops to identify:

- ☑ digital trends that will transform agriculture in the coming decade;
- ☑ digital capabilities an organization must have to be able to navigate these trends; and
- ☑ specific roles that public-interest actors such as CGIAR should claim in the evolving digital agriculture landscape.

This research strongly indicates that mission-driven "earth shots" in food security are achievable through digitally-enabled innovation and collective action, and a unified CGIAR can play a key role in achieving them. The products of this strategic research will inform the One CGIAR process in 2021 and beyond.

1 MODULE
## ORGANIZE PROGRESS

### 1.2.2 PROGRESS BY MODULES

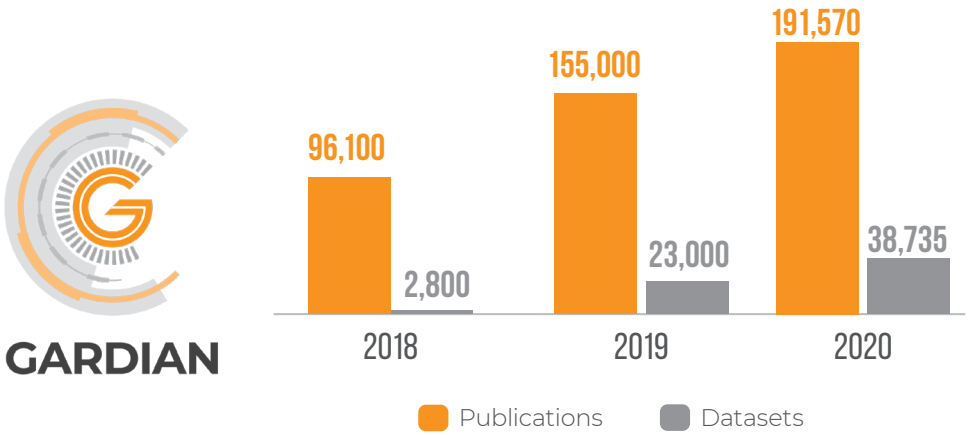
The Platform's Organize Module develops CGIAR culture and capacity to responsibly manage its research output, ensure its intellectual assets are open and FAIR, and to actualize the value of well-managed data. The Module's user-friendly tools help researchers collect and annotate their data and to realize the full value of well-managed data through enabling new analyses. The Module team helped update the CGIAR Open Access/Open Data Policy and advocated for the FAIR principles to be included in CGIAR performance and evaluation indicators. The team offered a vibrant array of workshops, webinars, and courses designed to build CGIAR's data management capabilities. GARDIAN features tools that support data managers in complying with open, FAIR and responsible data management. GARDIAN allowed researchers to not only find agricultural data, but also to fulfill its value by both working collaboratively and leveraging the analytical pipelines developed using the CG Labs analytic workbench. Helping researchers generate and use well-managed data is progressively building the value of CGIAR data and creating a culture of data stewardship that will accelerate food security research.

## DETAILED ANNEX

The GARDIAN data ecosystem offers technical tools designed to help researchers and data managers make CGIAR's data assets open and FAIR at every stage in the data management lifecycle. To facilitate participation in this one-stop knowledge base and answer "How do I do this?", GARDIAN made a data management toolkit available in 2020 with easy-to-use resources that allow users to more easily make their data assets open, FAIR, and responsibly managed. The toolkit includes links to data standards (e.g., the CG Core Metadata Schema, ontologies) as well as guides on responsible and FAIR data management plus workflows to easily make research outputs FAIRer. These tools help researchers and data managers more easily contribute to an increasingly open,

interoperable and reusable data pool that can be leveraged to quickly and responsively develop insights that can help transform agriculture. Additionally, the Agronomy Field Information Management System (AgroFIMS) enables data digitally collected in the field to be tied to data standards at the time of collection, making it "born-FAIR" data. Development efforts on AgroFIMS in 2020 included a streamlined interface and other options. Its developers will continue to respond to user needs and demand for this tool through the Excellence in Agronomy Initiative (EiA).

GARDIAN NOW POINTS TO OVER **190,000** PUBLICATIONS AND **38,000** DATASETS FROM SEVERAL PARTNERS ALONGSIDE ALL CGIAR CENTERS




The heart of the ecosystem is the GARDIAN discovery portal, which enables all CGIAR data and publication repositories to be searched along with those managed by a growing group of strategic partners including public agricultural research agencies, development funders, and the World Bank. In 2020, GARDIAN was used as a source of CGIAR data and/or publications by the FAO's Agris which provides access to publications for a wide range of partners; and by the GENDER Platform and the Alliance of Bioversity and CIAT (Alliance) as an enhancement to their webpages. Each new partner that joined the network to share data—be it an external strategic partner or a CGIAR program—further reinforces CGIAR's role as a trusted, capable broker or intermediary of open and FAIR agricultural data.

The Collaborative GARDIAN Labs (CG Labs) analytical environment is critical to this goal. In 2020, it increased the number of large datasets, computational power, and crop simulation model pipelines offered to users. These enhancements enabled more "plug-and-play" functions similar to those biomedical researchers have enjoyed for some decades and that have helped transform medical therapies. CG Labs allows researchers to not only find data via GARDIAN, but also to save it through CG Labs, upload data from other sources, and couple that with large climate or weather data that is part of the CG Labs bundled services. Thanks to a collaboration with the University of Florida's Agricultural Model Intercomparison and Improvement Project (AgMIP), data can be input to crop models commonly used by CGIAR and other agricultural researchers (e.g., DSSAT, WOFOST, EcoCrop), then used in spatial analysis and the development of maps using any one of hundreds of available pre-packaged R packages. CG Labs also allows built-in interaction with programming scripts that can be easily shared and made available for all to use. It was used in 2020 by members of CGIAR's Community for Spatial Information, Climate Change, Agriculture and Food Security (CCAFS), the Excellence in Agronomy Initiative (EiA) and other CGIAR and World Bank researchers for high compute, data-intensive tasks.

## DETAILED ANNEX

The Organize Module also leveraged semantic web technology to enhance the CGIAR Expert Finder (alpha version), which showcases CGIAR's research expertise and global reach. The Expert Finder allows CGIAR leadership and researchers, funders, and other stakeholders to quickly find who works where across the System (see Figure 1) based on keyword searches (e.g., food security, mechanization). It also allows researchers to easily find new collaborators on a given topic and/or in particular geographic areas or Centers, and it enables Center leads to easily get an interactive view of entities working around a particular theme (e.g., "aflatoxin", "gender nutrition"). Expert Finder can also provide an interactive mapping of CGIAR capabilities around any specific research theme or across themes. The tool currently compiles publicly available data (primarily publications and datasets) on CGIAR researchers, but it could be greatly improved were individual researchers encouraged to edit any non-authoritative information in the system, such as research summaries or geographic areas of expertise. It was used in 2020 as a source of information on people involved with the GENDER Platform, and, as an easily sliced resource, is being considered by a number of Centers as a potential content source for their web pages.

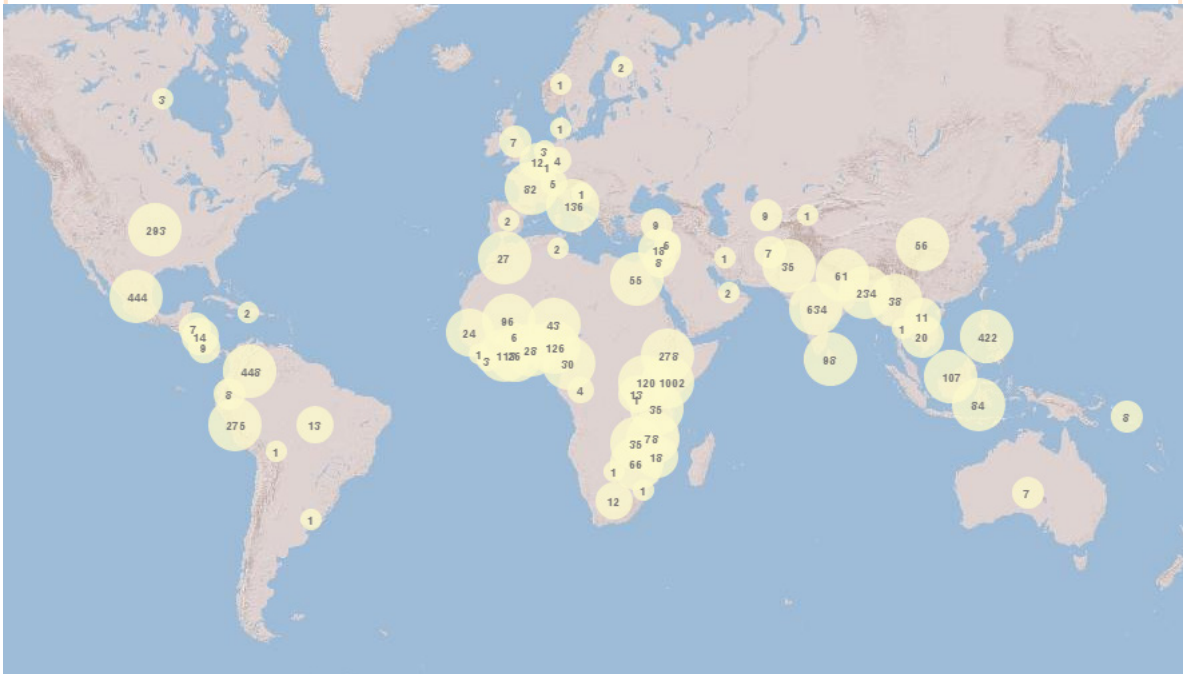



**Figure 1:** Landing page of Expert Finder (alpha) showing CGIAR researchers globally. Clicking on any yellow circle provides a list of people and their expertise areas, CGIAR affiliations, and other types of data.

### RELEVANCE TO COVID-19:

The Platform's Organize Module focuses on building culture, capacity, and tooling to help CGIAR researchers generate open and FAIR data assets that are responsibly managed. This work was not directly relevant to the Covid-19 Hub, but some of the tools developed through the Module have likely helped CGIAR researchers and their partners respond quickly and effectively to the pandemic. GARDIAN provides agricultural scientists, policy makers and analysts with a one-stop data and tool shop, for example, that can help in the development of agile regional or local pandemic responses. CG Labs, in particular, provides a collaborative environment and support for large computational analyses such as those required to quickly respond to changes in agricultural scenarios and outlooks wrought by Covid-19.

2 MODULE

# CONVENE PROGRESS

The Convene Module developed CGIAR thought leadership through BIG DATA's seven technical CoPs, organizing a virtual convention that reached millions of stakeholders, and forming new digital alliances. The CoPs grew to more than 5,000 members, becoming a key means by which CGIAR engages the wider sector to build digital research innovations. The Platform CoPs develop the program's technical depth; these released important community-driven products such as a tool to assist crop breeders in trait selection, a pipeline for delivering daily weather data to locations in the International Wheat Improvement Network, a household survey ontology, and a system for calculating GHG emissions from multiple crop events. These products were widely disseminated through 45 CoP-led webinars and workshops that reached more than 5,600 live attendees and an additional 11,700 viewers via YouTube channels. BIG DATA issued sector guidance on building enabling data ecosystems for startups and policy makers, technology standards for protecting agricultural data in information systems, and the potential risks of widespread use of AI in agricultural value chains.

## DETAILED ANNEX

BIG DATA was presented with a unique opportunity in 2020 to "walk the talk" on agile, adaptive, digitally-enabled collective action. For the Platform's fourth annual convention, it transitioned to be an inclusive, accessible, and fully online event. The event theme, Digital Dynamism for Adaptive Food Systems, examined food system resilience and highlighted how digital tools and technologies can help various food system actors sense, respond, and (re)build better systems in global food security crises.


The "socially-distanced" closing ceremony of the 2020 Convention featured eclectic, vibrant musical performances played to more than 500 staple crops. In the spirit of collective global action, the concert was a collaborative project with contributions by artists from across, Colombia, India and Kenya.

The 2020 Convention and recorded the greatest amount of CGIAR-wide engagement to date. Every CGIAR Center submitted an application to the Inspire Challenge and almost every Center was represented in the 10 finalist projects. With the Platform's support, CGIAR Centers submitted content that included 11 short videos offering a glimpse of digital dynamism at a Center level, as well 15 tech demos highlighting some of these centers' most cutting-edge digital innovations.

This was the most inclusive Convention to date. It provided an opportunity to showcase how each of the global Centers are employing digitally-enabled, dynamic methods to combat global food security challenges originating in current crises. The free virtual event attracted more than 1,300 active participants from around the world (75% of them external to CGIAR) who engaged in sessions and chats as well as on discussion boards around topics related to digital adaptation in global food security. The virtual format also enabled the Platform to engage a wider array of speakers. Our 150 speakers included both high-profile industry experts as well as digitally-enabled farmers from three regions. Half of these speakers were women.

The 2020 Youth In Data Workshop, which was conducted in parallel with the Convention, enrolled 130 youth from 49 countries. These young digital innovators interacted with and interviewed Convention experts and participants, resulting in them writing five blogs that were published on the BIG DATA website. They were key in the Platform's success in reaching more than 14 million people via the convention #BDPGLOBAL2020 tag across Twitter alone.

## RELEVANCE TO COVID-19:

COVID-19 has dramatically accelerated digital transformation in agricultural research for development. The global collaboration around open science, data sharing, and leveraging significant computational power to accelerate research in response to the COVID-19 crisis has given the agriculture development sector a glimpse of its own future. For the Platform, this was a breakthrough moment in which research and delivery became significantly more digital, such as in automating processes, increasing awareness of the value of good-quality data, demonstrating the increased importance of computational research as well as the agility and adaptability provided by digital services for marketing, extension, and distribution. It also precipitated the creation of two high-profile new alliances. CGIAR scientists joined X and Digital Green, Plantix, and One Acre Fund in developing a national-scale, dynamic, remote-sensing- based model for detecting maize crop emergence, unlocking food security monitoring with the timeliness and scale needed for managing future crises. COVID-19 lockdowns provided a unique opportunity to detect gas emissions patterns, and CGIAR researchers and HPE used its supercomputing capacity to develop a remote sensing model for detecting levels of nitrogen dioxide ($NO_2$) emissions from space, creating valuable new data assets needed to accurately model the interactions between economic activity and climate. HPE offered all CGIAR Science Leaders with in-kind use of these supercomputing resources.

# 2020 CONVENTION

**1ST** CGIAR
**ONE CGIAR**
CONVENTION

**14M**
REACHED ON TWITTER
#BDPGLOBAL2020

**2600+**
REGISTRATIONS

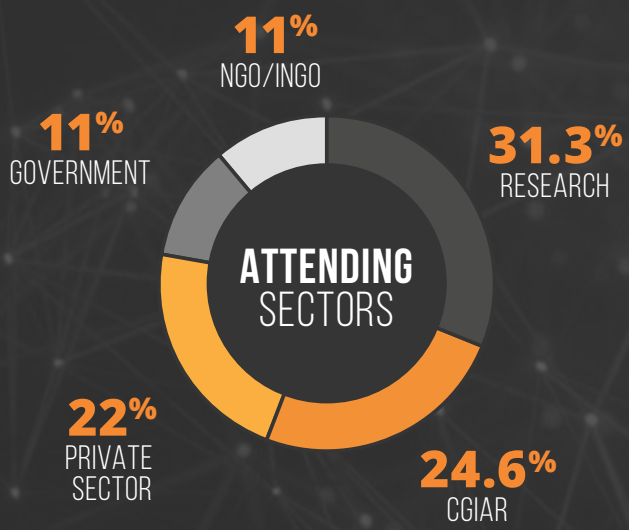
**1300+**
ACTIVE PARTICIPANTS

| | |
|---|---|
| 911 | ASIA |
| 788 | AFRICA |
| 465 | EUROPE |
| 386 | LATIN AMERICA |
| 419 | US AND CANADA |
| 44 | MIDDLE EAST |
| 50 | OCEANIA |

**150**
SPEAKERS

**25%**
YOUTH ATTENDEES

**50%**
FEMALE SPEAKERS




**700,000**
USD AWARDED

## DIGITAL DYNAMISM FOR ADAPTIVE FOOD SYSTEMS

3 MODULE

## INSPIRE PROGRESS

The Inspire Module continues to spark new digital innovation partnerships and overall engagement with the Platform, including its growing and vibrant technical CoPs (under Module Two).

In 2020, the Platform awarded a total of US$700,000 to seven innovative start-up projects using big data approaches to advance agricultural research and development. Additionally, BIG DATA awarded three, previously established, Inspire Challenge projects a total of US$100,000 in Rapid Response Grants—funding made available for agile, big-data-enabled projects working to tackle food system challenges related to the COVID-19 pandemic. The In 2020, several bilateral funders expressed an interest in the Challenge; two of them investing a total of US$500,000.

In total, these 10 winning teams included collaborations between a total of nine CGIAR Centers and Research Programs and 22 diverse external partners, including start-ups, governmental agencies, universities, and private sector businesses. The Inspire Challenge has begun to field test innovations that demonstrate new digital pathways to impact for CGIAR research and foster whole system agility and adaptation through the use of digital tools.

## DETAILED ANNEX

Some important results Inspire Challenge projects achieved in 2020 include:

PlantVillage Nuru played a key role in responding to the 2020 locust swarms in East Africa by adapting its groundbreaking, AI-informed approach to pest and disease monitoring. Borrowing from the Plantvillage Nuru application blueprint, the team developed a new smartphone application in just one month for monitoring locusts across the region. The 240,000+ locust records captured via the eLocust3m app enabled countries to systematically act to control the swarms. Since February 2020, the FAO estimates that this effort in East Africa has averted the loss of commercial agricultural products valued at US$1.5 billion—saving the livelihoods of 34 million people.

Hungry Cities leveraged its high frequency data on the metropolitan Nairobi food system to get fresh fruit and vegetables (FFV) to the people who needed them the most during the COVID-19 pandemic. In June 2020, the project's data showed that 90% of households in the Nairobi slums were experiencing dire food insecurity, including reduced FFV consumption, primarily due to reduced household incomes. The project's data-driven insights led Twiga Foods, a leading agricultural produce supplier in Kenya, to reduce its FFV prices in its markets that serve hundreds of thousands of low-income consumers.

Another project has become an instrumental partner to the General Statistics Office of Vietnam in assessing COVID-19-related food security shocks at a national level. Since 2019, the team has continually monitored informal food flows in five traditional markets in Hanoi, creating a valuable dataset of more than 200 million total data points related to the impacts of and actor behavior changes before, during, and after the first and second waves of COVID-19. The early analysis reveals a sharp decrease in the average number of people seen daily in the urban markets in March 2020, compared to 2019 baseline numbers, and 38% of previously frequent visitors—those who had typically visited at least five times a week pre-COVID—came less often or totally avoided the markets months after the pandemic's onset.

## RELEVANCE TO COVID-19

The COVID-19 crisis presented BIG DATA with the opportunity to demonstrate that digital tools, methods, and partnerships provide the critical agility and adaptability needed to manage food security shocks. As a result, the Platform realigned some internal budgets to run the Rapid Response grant challenge in which previous awardees could seek additional funding to leverage their existing projects and partnerships to implement response, recovery, and resilience interventions for mitigating local food security impacts.

The awards and their projected impacts are as follows:

- Rapid diagnostics of COVID-19 farming impacts developed a mobile phone-based panel survey approach to assess COVID-19 impacts on farming, which is being tested at scale across Kenya through the "Let it Rain" campaign. In 2020, the team completed two surveys with more than 2,000 farmers on COVID-19 impacts. Two additional surveys will be conducted in 2021. The complete set of findings will inform targeted, digital advisories for farmers and programmatic responses through info briefs.
- Herd opportunity will deploy an interactive digital course to train livestock farmers on how to prevent the spread of infectious diseases, including COVID-19, on and off the farm. The team will pilot the course in Kenya and aims to reach between 5,000 and 10,000 livestock farmers within the first four months. Within two years, the team aims to scale up its training to more than 250,000 farmers across sub-Saharan Africa. Using data from the interactive course, the project will also generate live data on hygiene practices and infection risk among livestock farmers through an open access data dashboard.
- Eyes on the ground for agricultural microcredit aims to unlock agricultural microcredit to farmers in Odisha, India, by enabling a microfinance lender to monitor the loans its makes using satellite and smartphone imagery. The team aims to pilot this approach by issuing US$400 loans to 450 farmers in 50 randomly-selected villages.

# INSPIRE 2020

**120** APPLICATIONS

**3** RAPID RESPONSE GRANT WINNERS

**7** INSPIRE PILOT WINNERS

USD$ **800,000** TOTAL AWARED

TO INSPIRE CHALLENGE PILOT PROJECTS AND RAPID RESPONSE GRANTS COMBINED

**21** ACTIVE PROJECTS

USD$ **500,000** IN NEW EXTERNAL FUNDING

INSPIRE CHALLENGE & RAPID RESPONSE GRANT PARTNERSHIPS

**9** CGIAR CENTERS AND RESEARCH PROGRAMS

**22** EXTERNAL PARTNERS

Inspire Challenge

## 1.2.3 VARIANCE FROM PLANNED PROGRAM FOR THIS YEAR

**A. Have any promising research areas been significantly expanded?**

The COVID-19 crisis prompted the re-alignment of some internal budgets to issue "Rapid Response" digital innovation grants to demonstrate that the agility and adaptability enabled by deploying digital tools and approaches would be important for response, recovery and resilience in future food security shocks. The Platform's strategic research on CGIAR roles in the digital agriculture landscape has generated several information products that are informing the One CGIAR reorganization. The BIG DATA team's ongoing work focused on building the Digital One CGIAR is expanding beyond what was originally envisioned.

**B. Have any research lines been dropped or significantly cut back?**

The Platform evaluated external shared services and found that the value-for-money and ongoing costs of most external subscriptions not attractive as an approach for building enduring CGIAR capabilities. As a result, the shared services focus shifted toward providing and supporting CG Labs, an internal analytic workbench and tool for CGIAR researchers.

**C. Have any Flagships or specific research areas changed direction?**

No Modules or research areas have changed the direction outlined in the original program design.

## 1.2.4 ALTMETRIC AND PUBLICATION HIGHLIGHTS

The Platform outcomes are focused on organizational change and strategy rather than scientific publications. The eight peer-reviewed publications generated by the Platform are reported under their specific projects.

The Platform showed success in building its brand and establishing CGIAR as an authority in the digital agriculture space. The BIG DATA website received approximately unique 200,000 visits from 65,000 users in 2020, increases of 65% and 51%, respectively, with sessions increasing 44%. Our social media audience (Facebook, LinkedIn, Instagram, YouTube, and Twitter) doubled, growing 97% to a total of 20.6K followers and subscribers. We published a total of 93 newsletters that recorded an open-rate of 3% greater than the industry average.

Through coordinating with the CGIAR System Office, the Platform hosted the first One CGIAR-branded convention and recorded the greatest CGIAR-wide engagement to date. With the Platform's support, CGIAR Centers submitted 11 short videos and 15 tech demos highlighting Digital Dynamism across CGIAR.

The BIG DATA communications team supported our technical CoPs to host 43 webinars – up 400% from 2019 -- and to publish 61 newsletters and 69 blog posts, increases of 280% and 53%, respectively. The sum of all CoP members reached grew to more than 5,000 by the end of 2020.

In response to COVID-19, the Platform launched a digital discussion series: Big data solutions to COVID-19 impacts on food security. In the course of seven episodes, we examined on-the-ground realities, emerging digital solutions, and big data opportunities that have arisen from the pandemic. The series featured 22 speakers from organizations such as start-ups and players in the agtech private sector as well as organizations such as IFAD and GSMA. It attracted 4,300 viewers in total. Several participating organizations appeared as Convention speakers (IFAD, GSMA, Digital Green,

MoooFarm). This provided an excellent opportunity for BIG DATA to showcase how the Rapid Response grants and Inspire Challenge projects responded to the COVID-19 crisis. We published seven session-summary blogs and a series summary that listed the challenges and opportunities brought about by the pandemic.

## 1.3 CROSS-CUTTING DIMENSIONS (AT PLATFORM LEVEL)

### 1.3.1 GENDER

**A. List any important Platform research findings**

In 2020, the Platform team developed, adopted, and implemented its first strategy for advancing gender equity in agricultural research and development. With input from the Generating Evidence and New Directions for Equitable Results (GENDER) Platform, BIG DATA designed a four-pronged strategy for improving our understanding of relationships among gender, agriculture, and rapidly digitizing economies and developing FAIR data systems that enable gender data to be used to its full potential. The strategy was confirmed by the Platform Steering Committee in May 2020.

Collaborative research between BIG DATA and the GENDER Platform centered on the development of a novel, timely, large-scale assessment of women's economic empowerment was completed and published in 2020. This approach leveraged many billions of data points generated by mobile phone usage patterns (e.g. number of calls, location, movement of phones) in Uganda that were anonymized. The team reviewed literature on the uses of these call records for establishing credible proxy measurements for socioeconomic or demographic data, selected some indicators inspired by the literature and body of research into women's economic empowerment, commissioned a 10,000-person phone survey related to those indicators, and used the survey to interpret patterns in the larger dataset at a national scale. The method showed promise as a potential enhancement to more-established methods of gender research, particularly in terms of the large scale and timeliness of data and phone-based methods.

BIG DATA and GENDER also co-funded an extensive review of human-centered design as it intersects with social inclusion and digital agriculture. The full report will be disseminated in 2021.

In line with its strategic objective to provide funding and support for gender-sensitive agricultural research and development projects, the Platform added rigorous gender mainstreaming components to the 2020 Inspire Challenge and Rapid Response Grants. Each proposal was required to elaborate on the potential gendered implications of their project and the strategies that would be employed to manage them. Furthermore, applicants were required to indicate the gender breakdown of the principle team members. Each of these gender-related components were weighted in the Inspire Challenge and Rapid Response judging rubrics. Using these criteria, the Platform awarded a total of US$800,000 to 11 total gender-sensitive agricultural research and development projects (four Rapid Response, seven Inspire Challenge) in 2020. Each project will be reviewed for insights related to gender and big data in agricultural innovation at their one-year mark of maturation and the relevant learning shared with the wider agricultural innovation sector.

Finally, the Platform hosted a gender track at the 2020 virtual CGIAR Convention on Big Data in Agriculture. The track featured three sessions that took deep dives into big data topics relating to the empowerment of women in agriculture. Thirteen speakers from 12 key public and private organizations presented alongside CGIAR researchers on topics ranging from gender data research protocols to non-traditional big data sources for cultivating women's empowerment.

**B. What have you learned? What are you doing differently?**

Call-record based methods need to be anchored in a deep (and well-researched) understanding of the context in which they are used, and must be significantly ground-truthed. That said, they show promise as a useful new method for socioeconomic research. The team is looking to pilot this approach in other countries in which the call-record based approach can be further refined and validated.

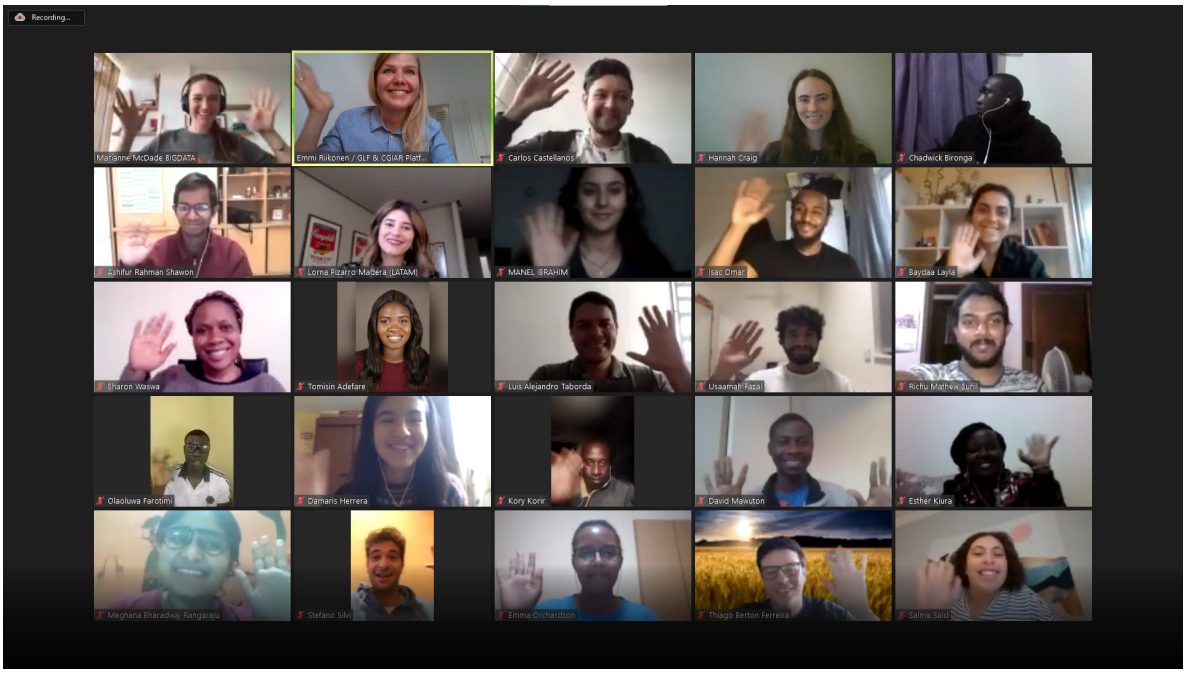
The BIG DATA and GENDER review of digital social inclusion in now a key input for the design of a proposed new cross-cutting digital initiative for One CGIAR.

**C. Have any problems arisen in relation to gender issues or integrating gender into the PTF's research?**

The Platform is not a research Platform and only ventures into research in situations in which it could develop an important new digital capability for CGIAR. That said, the Platform integrated gender into its innovation process, data standards work, and some dedicated new research without any difficulty.

## 1.3.2 YOUTH AND OTHER ASPECTS OF SOCIAL INCLUSION / "LEAVING NO-ONE BEHIND"

Transitioning the annual convention to a virtual event enabled the organization to engage with a very diverse and far-ranging group of people. We estimate that, of the several thousand registrants and 800-1,000 active participants, some 25% were youth. One key contributor to this successful outreach was the 2020 Youth in Data Digital Workshop led by the Platform communications team.




The 2020 Youth In Data Workshop call received 600 applications -- a 650% increase compared to the previous year. A total of 130 youth from 49 countries were enrolled.

The Youth in Data Workshop initiative, which was created in 2018 to engage with young digital innovators, introduces participants to important themes around digital agriculture and how using big data approaches to agricultural development can accelerate food security goals. Delegates were also trained in how to use digital media as a powerful tool to communicate about development. They learned the basics of social media, interviewing, blog writing, data reporting, and other media skills.

The 2020 Youth In Data Workshop call received 600 applications -- a 650% increase compared to the previous year. A total of 130 youth from 49 countries were enrolled. These young digital innovators interacted with and interviewed experts and participants at the Convention, and were key in the Platform's success in reaching more than 14 million people with the convention #BDPGLOBAL2020 tag across Twitter alone.

**A. List any important Platform research findings**

BIG DATA is not a research Platform, rather it seeks to be a broker of new partnerships and innovations for impact, and to serve as a mechanism for remaining abreast of and contributing to trends in the rapidly evolving digital agriculture space. Based on the Platform's experience in running a virtual Convention, we found meeting virtually is a significantly more youth inclusive approach compared to conventional, in-person events. Any future events should have a significant digital/virtual component as a means of building global engagement around the digital CGIAR and to provide young researchers and students the opportunity to participate without expending the additional time and money needed for attending a conference in person.

**B. What have you learned? What are you doing differently?**

The Platform or any of its potential follow-on activities under One CGIAR should continue to develop digital communications channels and virtual events to be able to more effectively reach youth.

**C. Have any problems arisen in relation to youth issues or integrating youth into the Platform research?**

BIG DATA is not a research Platform. That said, it has encountered minimal problems with building youth engagement around digital agriculture and CGIAR.

## 1.3.3 CAPACITY DEVELOPMENT

The Platform works towards milestones and outcomes intended to help build CGIAR's digital capabilities and new digital pathways to impact for agricultural research. In 2020, the Platform piloted a novel data science academy with a cohort of 25 CGIAR scientists, using the popular and accredited online learning service Coursera. In addition, the Platform continued to offer an online course in FAIR data standards for agricultural research. This content was further customized and delivered specifically as it relates to the collection, curation, storage, and re-use of agronomic data.

Virtual engagement with the Platform CoPs grew significantly in 2020, with our technical CoP hosting 43 online seminars—up 400% from 2019. The topics included advancing technical collaboration and best practices, data management, scaling digital farming innovations, and integrating digital survey tools into research.

Partnerships are a key means by which organizations can access new capabilities and integrate these into how they do business. Across all Modules, the Platform formed or brokered a total of 72 partnerships linking CGIAR to new capabilities. This includes the partnership of nine CGIAR Centers with 22 external actors through the seven Inspire Challenge projects. The Inspire Challenge project "Let it rain" developed new content for the extremely popular agricultural advisory TV program Shamba Shape-up, with a viewership of around 6,000,000.

## 1.3.4 CLIMATE CHANGE

During 2020, the Platform brokered some important new collaborations that build cross-cutting, unified digital capabilities for agroclimatic characterization and forecasting, such as the development of new tools to help crop breeders in trait selection. A CCAFS team used CG Labs, the BIG DATA analytic environment, for example, to develop climate scenarios for several crops and value chains in sub-Saharan Africa and Vietnam. An external partner using CGIAR data was subsequently able to use CG Labs to assess the profitability of fertilizer use on a sub-continental scale in Africa, providing critical context for informing agroecological interventions.

02

# EFFECTIVENESS AND EFFICIENCY

## 2.1 MANAGEMENT AND GOVERNANCE

No changes to overall Platform governance and management were made in 2020. The Platform still has three active governance bodies: the Management Team, which consists of leaders from the technical CoPs and members of the Secretariat; the Steering Committee, which includes representatives of CGIAR Centers and Programs, as well as and external experts from the private sector, funding organizations, and research; and an entirely external international advisory board.

## 2.2 PARTNERSHIPS

### 2.2.1. HIGHLIGHTS OF EXTERNAL PARTNERSHIPS

In 2020, the Platform was able to build two new alliances with global firms to push the discipline of digital agriculture forward: X and HPE. With X, BIG DATA brokered data sharing and new model development efforts to implement in-season, national scale modeling to detect maize crop onset in Kenya—building important new capabilities for in-season food security monitoring. With HPE, CGIAR researchers used HPE's Superdome Flex supercomputing environment to train a model for monitoring nitrogen dioxide—a chemical precursor of atmospheric ozone— emissions from space. The Platform experimented with running open data science competitions using an online competition platform to invite "citizen data scientists" to work on analytic problems with CGIAR data. This attracted participants from more 80 countries and proved to be an excellent means of identifying talented new partners.

Under Module Three, the Platform provided and widely promoted a partner-matching service in support of the Inspire Challenge. Its application requires teams consisting of CGIAR and non-CGIAR members. This service facilitated the creation of 130 teams that were able to apply.

### 2.2.2. CROSS-CGIAR PARTNERSHIPS

The Platform was able to form or broker more than 70 digital partnerships supporting CGIAR research innovations and develop new digital pathways for delivery and impact. Twelve of the 13 CGIAR Centers participated in one or more of these alliances. There is no other unit in CGIAR with the explicit, specific mandate to develop cross-cutting digital partnerships for CGIAR. Cross-cutting digital partnerships will be a critical capability that will continue to be developed under One CGIAR.

## 2.3. INTELLECTUAL ASSETS

**Have any intellectual assets been strategically managed by the PTF (together with the relevant Center) this year?**

The Platform led revision of the 2013 CGIAR Open Access and Data Management Policy, which should be finalized in 2021, and worked with the Information and Data Managers CoP to implement an updated metadata schema for CGIAR data. Both efforts were guided by the FAIR Data Principles. The Platform also entered into partnerships with digital companies in the pre-competitive space, where there is shared interest in advancing digital methods or developing data and technology standards for the open, public-good science infrastructure that would also be of potential commercial interest to the partners. The Platform retained an intellectual property lawyer with specific expertise in biodiversity and plant genetic resources to help guide these efforts.

**Indicate any published patents and/or plant variety right applications (or equivalent)**
No patents were filed.

**List any critical issues or challenges encountered in the management of intellectual assets**
There are significant nuances to be navigated to find the aligned interests across in developing public, private, non-profit organizations. The Platform has distilled its partnership development experience into broad recommended collaboration principles that its partners are requested to adhere to. This appears to be a promising way to help formal bilateral partnerships mature into multi-stakeholder alliances around shared interests.

## 2.4 MONITORING, EVALUATION, IMPACT ASSESSMENT AND LEARNING (MELIA)

The Platform designed and implemented a robust approach for wide-ranging stakeholder consultation for development of a CGIAR Digital Strategy through interactive focus group brainstorming, one-on-one interviews, and large, Center-based discussion groups to identify key capabilities CGIAR will need to maintain or further develop to fully leverage data and digital technologies to implement the 2030 Research and Innovation Strategy. The Digital Strategy is expected to be finalized and released in 2021.

Under Module Two, the Crop Modeling CoP assessed the state of modeling in light of climate change and made specific recommendations on how to proceed in resource and data-constrained settings. As part of the overall digital strategy research conducted by the Platform, the team assessed CGIAR's digital capabilities and the roles the System should be claiming in the overall digital agriculture landscape.

An external evaluation of the Inspire Challenge was conducted in 2020, looking at its efficacy in targeting innovation and brokering effective new partnerships for research and research delivery.

## 2.5 EFFICIENCY

### Organize

The Organize Module continued to develop key organizational systems and data standards to support CGIAR efficiency. GARDIAN pointed to more than 190,000 publications and 38,000 datasets, respective annual increases of 22.5% and 65%, from several partners as well as all CGIAR Centers. The CG Labs secure analytic environment in the GARDIAN ecosystem, which enables researchers to find data and collaborate on analyses, was enhanced and used to develop high-value products including climate risk profiles for several countries. CG Labs enabled CGIAR researchers to assemble CGIAR and partner data and collaborate on using common, open, high-compute analytic pipelines as well as visualization and mapping tools. This will be a key capability to develop as new research initiatives are rolled out under One CGIAR.

### Convene

The Convene Module made significant strides in developing CGIAR's digital partner networks. Seven CoPs increased the number of seminars and engagements they held, growing CoP membership to 5,000 CGIAR and non-CGIAR participants. The CoPs released important community-driven products for the improved efficiency and efficacy of research, including guidelines for minimum data standards for crop modeling, approaches to scaling digital agriculture interventions, and implementing cloud-based, harmonized approaches to socioeconomic surveys.

### Inspire

The Platform provided seven start-up grants to data-driven partnerships through the Inspire Challenge. Its growing portfolio, which now has 21 projects, demonstrated new impact and delivery pathways for CGIAR research.

## 2.6 MANAGEMENT OF RISKS TO YOUR PLATFORM

Contextual risks: In response to the COVID-19 crisis, BIG DATA had to radically, and rapidly, adapt how it conducted business. Most of the programmatic objectives were met and some were even exceeded due to an increased appreciation for digital capabilities worldwide. Some access to the field was constrained during lockdowns, which caused some programmatic delays.

Programmatic risks: The Platform radically, and successfully, redesigned its program implementation to be virtual and digital, achieving most programmatic targets for the year. One notable exception concerns the Inspire Challenge grant projects. Access to the field was constrained, causing delays in several projects. A handful of these projects were issued a no-cost extension and urged to wrap up their projects by the end of Q1 2021.

Institutional risks: The Platform faced no significant institutional risks in 2020.

## 2.7 USE OF W1-2 FUNDING

One of the underlying assumptions of the Theory of Change for the Platform is that data, digital tools, and digital capabilities can be used to build needed dynamism and much-needed agile adaptability in agri-food systems. The COVID-19 crisis resulted in less of the budget being spent on travel and in-person events, which enabled the Platform to re-align its funds to develop and run "Rapid Response" grants under the Inspire Challenge. All previous Inspire Challenge grant awardees were invited to apply for additional funding that would leverage the on-the-ground digital projects to address COVID-19 response, recovery, and resilience in their specific project contexts. The Platform Steering Committee approved this budget re-alignment and the Platform was able to quickly put its assumptions about digital dynamism to the test. The small grants awarded supported the rapid reconfiguration of a picture-based insurance project to provide data-driven micro-credits to vulnerable farmers, use of a social media training platform to rapidly deploy engaging COVID-19 adaptation content, and in-season assessment of food production or distribution disruptions.

03

# FINANCIAL **SUMMARY**

The Platform has navigated an uncertain funding year and an array of programmatic and context risks, remaining in-budget while fulfilling virtually all of its programmatic objectives from the original proposal and the original plan for 2020. The Platform is on track to expend all its W1/W2 funds and reach its milestones by the end of the 2021 calendar year.

# LIST OF TABLES

## TABLE 2: CONDENSED LIST OF POLICY CONTRIBUTIONS IN THIS REPORTING YEAR (SPHERE OF INFLUENCE)

| Title of policy, legal instrument, investment or curriculum to which CGIAR contributed | Description of policy, legal instrument, investment or curriculum to which CGIAR contributed | Level of Maturity | Link to sub-IDOs | CGIAR cross-cutting marker score<br>G = Gender<br>Y = Youth<br>CD = CapDev<br>CC = CC | | | | Link to OICR<br>(obligatory if Level of Maturity is 2 or 3) or link to evidence (e.g. PDF generated from MIS) |
|---|---|---|---|---|---|---|---|---|
| | | | | G | Y | CD | CC | |
| 632 - National Fisheries Strategy | | Level 1 | • Enhanced capacity to address climatic risks and extremes (Mitigation and adaptation achieved.).<br>• Enhanced institutional capacity of partner research organizations<br>• Improved forecasting of impacts of climate change and targeted technology development. | 0 - Not Targeted | 0 - Not Targeted | 2 - Principal | 2 - Principal | OICR3874 |
| 709 - Technology standards for operationalizing agricultural data protection | The Platform released sector guidance mapping common data protection policies (e.g., the General Data Protection Regulation) and ethical frameworks (e.g., BIG DATA's responsible data guidelines) to an applicable framework for assessing risk and the relevant formal technology standards (e.g., NIST, ISO) for implementing a data protection infrastructure. | Level 1 | • Conducive agricultural policy environment | ? - Too early to tell | ? - Too early to tell | 1 - Significant | ? - Too early to tell | Technology standards for agricultural data protection report published and disseminated at BIG DATA Convention. |

## TABLE 3: LIST OF OUTCOME/ IMPACT CASE REPORTS FROM THIS REPORTING YEAR (SPHERE OF INFLUENCE)

| Title of Outcome/ Impact Case Report (OICR) | Link to full OICR | Maturity level |
|---|---|---|
| OICR3835 - "Let it Rain" game led to 25,312 successful players and new sign ups to iShamba | Link | Level 2 |
| OICR3874 - Use of PeskAAS led to improved fisheries management | Link | Level 3 |
| OICR3877 - General Statistics Office of Viet Nam explored the use of free Wi-Fi to measure COVID impact on traditional markets | Link | Level 1 |
| OICR3883 – More than 250,000 users reached—with 62,500 users and 41 organizations active—on the new web platform | Link | Level 2 |
| OICR3912 - Picture-based crop monitoring is adopted by six institutions to improve their agricultural insurance and financing services | Link | Level 2 |

## TABLE 4: CONDENSED LIST OF INNOVATIONS BY STAGE FOR THIS REPORTING YEAR

| Title of innovation with link | Innovation Type | Stage of innovation | Geographic scope (with location) |
|---|---|---|---|
| 724 - CGIAR Core Metadata Schema version 2 | Research and Communication Methodologies and Tools | Stage 3: available/ ready for uptake (AV) | Global |
| 1102 - SEONT: Agricultural Household Survey Ontology | Research and Communication Methodologies and Tools | Stage 2: successful piloting (PIL - end of piloting phase) | Global |
| 1259 - Gap analysis of crop modeling activities within CGIAR and collaborators | Research and Communication Methodologies and Tools | Stage 1: discovery/proof of concept (PC - end of research phase) | Global |
| 1305 - Innovative Risk Management Tools for the Poor | Research and Communication Methodologies and Tools | Stage 4: uptake by next user (USE) | Multi-national, Ethiopia, Kenya |
| 1702 - Auto-CCAFS-MOT: a batch calculator tool for farmers' GHG emissions | Production systems and Management practices | Stage 3: available/ ready for uptake (AV) | Global |
| 1706 - Let it Rain Game: The game and linked iShamba service is the innovation | Research and Communication Methodologies and Tools | Stage 2: successful piloting (PIL - end of piloting phase) | National, Kenya |
| 1709 - Rappseq: A cloud platform for identifying bacterial pathogens from long-read DNA sequence data | Other | Stage 1: discovery/proof of concept (PC - end of research phase) | Global |
| 1712 - Consideration of remote sensing to facilitate minimum data set collection | Production Systems and Management Practices | Stage 1: discovery/proof of concept (PC - end of research phase) | Global |
| 1713 - Target Population of Environment (TPE) modelling as a synergy between BDP and EiB platforms | Research and Communication Methodologies and Tools | Stage 2: successful piloting (PIL - end of piloting phase) | Global |
| 1728 - GARDIAN Labs, an open-access service enabling researchers to collaborate and share data and analytic approaches and tools | Research and Communication Methodologies and Tools | Stage 2: successful piloting (PIL - end of piloting phase) | Global |
| 1738 - Potato AI model able to detect potato late and early blight | Biophysical Research | Stage 3: available/ ready for uptake (AV) | Global |
| 1766 - National scale, in-season crop modeling for monitoring food security shocks | Production Systems and Management Practices | Stage 1: discovery/proof of concept (PC - end of research phase) | National, Kenya |
| 1776 - Predicting yields using biophysical crop simulations, machine learning, and remote sensing | Production Systems and Management Practices | Stage 2: successful piloting (PIL - end of piloting phase) | Sub-national, India |
| 1989 - High-value data product demonstrating value of Open and FAIR data: Profitability of fertilizer use in Africa | Production Systems and Management Practices | Stage 2: successful piloting (PIL - end of piloting phase) | Regional, Sub-Saharan Africa |
| 1990 - High-value data product demonstrating value of Open and FAIR data: Assessing climate risk profiles and crop suitability | Biophysical Research | Stage 2: successful piloting (PIL - end of piloting phase) | Regional, Eastern Asia, Sub-Saharan Africa, Southern Asia |

## TABLE 5: SUMMARY OF STATUS OF PLANNED OUTCOMES AND MILESTONES (SPHERE OF INFLUENCE-CONTROL)

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M1 | M1 Outcome: 1.1. A demand-driven analytics environment is available. | · Enhanced institutional capacity of partner research organizations<br>· Increased capacity for innovation in partner research organizations<br>· Improved forecasting of the impacts of climate change and targeted technology development<br>· Increased capacity for innovation in partner development organizations and in poor and vulnerable communities | Two high-value data products were created using GARDIAN's data and CG Labs analytics environment: (1) Climate risk profiles for 16 countries using pre-loaded data, tools, and models in CG Labs; (2) Analysis of fertilizer profitability in Africa using more than 200 GARDIAN datasets. Machine learning was employed to create crop yield curves based on fertilizer inputs, and input/output price data was used to determine where fertilizer use was profitable. All R scripts developed to clean, process, and standardize datasets are available through GitHub and can be called from CG Labs for reuse or modification by any user. | 2020 - 1.1.7. 2020 - At least 2 high-value data products created through CoP engagement and by leveraging GARDIAN and the analytics made available through GARDIAN Labs. | Complete | Climate risk profiles for 16 countries were developed by a CCAFS team of researchers using CG Labs for their high-computing needs and analysis. The results are currently available via Google Drive, while the team finalizes its reports and publications. The assessment of fertilizer profitability for Africa is complete and in review for publication in early 2021. The R scripts used to clean and process the more than 200 datasets used in this study are available, and will be available in association with individual datasets in GARDIAN in 2021. | Climate risk profiles for 16 countries: https://drive.google.com/drive/folders/10WrM-F_79Yg-2oU9P3jxzyypq-JsW9weki<br>Fertilizer profitability assessment –in review for publication in Nature Food; in response process: https://drive.google.com/file/d/1nop8kmLSgOcqYjHIIN9JJfLZxy2wIrqB/view?usp=sharing |
| | | | | 2020 - 1.1.8. 2020 - At least one cloud-based machine learning application was developed in conjunction with BIG DATA CoPs, and scientists from and beyond CGIAR. | Complete | The ML approaches used in developing climate risk profiles for 16 countries is in the process of being published. The assessment of fertilizer profitability for Africa is currently in review and should be available as a Nature Food publication in early 2021. | None at this time; will be available in early 2021. |
| | | | | 2020 - 1.1.9. 2020 - At least two new services operating on GARDIAN's data pool were created, including one to easily clean and process GARDIAN data, and test pipelines to simulation tools. | Complete | About 200 relevant GARDIAN datasets were cleaned, processed, and standardized, before a machine learning model could be used to generate insight on crop yield response to fertilizer application for different countries. | Fertilizer profitability assessment data cleaning and processing R scripts: https://github.com/reagro/fercrop |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M1 | | | | | | A unique R script was developed to clean, process, and standardize each dataset, with a "mother" script to run them all. All the scripts are available via GitHub and can be called from CG Labs for reuse or modification by any user. Pipelines to three mechanistic crop simulation models have also been made available via CG Labs for easier leveraging: WOFOST, EcoCrop, and DSSAT. | |
| | M1 Outcome: 1.2. CGIAR resources are discoverable and reused. | · Enhanced institutional capacity of partner research organizations<br>· Increased capacity for innovations in partner research organizations<br>· Increase capacity of beneficiaries to adopt research outputs<br>· Increased capacity for innovation in partner development organizations and in poor and vulnerable communities<br>· Improved capacity of women and young people to participate in decision-making | GARDIAN was launched in 2018 to promote the discoverability of CGIAR data and publications with tools for data visualization and back-end services to accommodate data privacy, ethics, and clarity on licensing. GARDIAN enhancements for 2020 include: CG Labs offering pre-loaded data and toolsets for analysis visualization and data insight; the launch of a tool for data query and aggregation (that is in the pilot stage); secure transfer and sharing of data; and the inclusion of a data management toolkit, including a FAIR data and publications upload workflows. GARDIAN also indexed data from new external data sources. | 2020 - 1.2.9. 2020 - GARDIAN includes data assets from at least four new partners' repositories. | Complete | GARDIAN's data discovery portal was expanded in 2020 to include data and publications from the World Bank's Living Standards Measurement Survey catalog, and data from the ISRIC Soil Grids database, offering geographic administrative boundary and user-based global visual mapping and download of several soil parameters. Data from the 2017 Spatial Production Allocation Model (SPAM) for crops grown in Africa was also indexed and made similarly visualizable. IRRI's Farm Household Survey Database (FHSD), with about 150 datasets, was also made discoverable via GARDIAN. | https://gardian.bigdata.cgiar.org |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M1 | | | | 2020 - 1.2.10. 2020 - New GARDIAN enhancements that include: "CG Labs" for analytical, visualization and data insight services and tools; data query and aggregator pilot; tested the Globus software as a key component for enabling the secure transfer and sharing of data; testing and refinement of FAIR data and publications upload workflows for partners, including machine-aided detection of personally-identifiable information (PII), metadata addition, and ontology annotation. | Complete | CG Labs was tested, refined, and made available in the GARDIAN data ecosystem. Users can set up an individual or team work space to collaborate on data processing and analytics, working together on data found through GARDIAN or elsewhere. Data can be securely exchanged and saved for collaborative processing and analysis via the CG Labs R and Python analytical environment. Tooling includes a workflow to add CG Core metadata, choose licenses, and add ontology terms to describe data, allowing data to easily be made FAIR. PII-checking is automated using this workflow. | CG Labs: https://gardian.bigdata.cgiar.org/labs.php<br>CG Labs use case demos: https://www.slideshare.net/cgiarbigdata/2020-big-data-cg-labs-use-cases |
| | M1 Outcome: 1.3. Standards and semantics are utilized to enable FAIR agricultural data. | · Increased capacity for innovations in partner research organizations<br>· Increased capacity for innovation in partner development organizations and in poor and vulnerable communities | Webinars and training sessions on the CG Core Metadata Schema v.2.0 were organized. Meetings with Gender Platform representatives and CGIAR Metadata Working Group members ensured gender-disaggregated and other data assets were as FAIR as possible. New data collection apps were added to AgroFIMS: KDSmart, Open Data Kit (ODK) and Field Book. The latter are Android-based. | 2020 - 1.3.11. 2020 - Continued support for CG Core Metadata Schema v.2.0 refinement and implementation across Center publications and data repositories was supported user-friendly workflows and tools, training, and documentation. | Complete | Webinars and training sessions on the CG Core Metadata Schema v.2.0 were organized and consultations with GENDER Platform representatives and CGIAR Metadata Working Group members held to ensure that gender-disaggregated and other data assets were as FAIR as possible. A session—"Gender data: knowledge gaps, research protocols, and improving data systems"—held during the Big Data Platform's 2020 Annual Convention highlighted the importance of using metadata standards and rich annotation to make data more FAIR. | https://bigdata.cgiar.org/virtual-convention-2020/schedule/ |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M1 | | · Increase capacity of beneficiaries to adopt research outputs<br>· Improved forecasting of impacts of climate change and targeted technology development | | 2020 extended to 2021 - 1.3.12. 2020 - Further development of AgroFIMS to pilot: Digital collection of "non-traditional" agronomic survey and demonstration data; use of AgroFIMS on ODK and KDSmart. | Extended | Two new data collection apps were added to the AgroFIMS repertoire in addition to KDSmart: ODK and Field Book, both of them Android-based. Progress was made in developing AgroFIMS to enable data collection from agronomic survey-type trials through design and mockups; however, this work will not be developed until users from the Excellence in Agronomy Initiative can specify requirements and co-create the tool in 2021 for maximizing success. | https://agrofims.org/about |
| | M1 Outcome: 1.4. Enhance capacity, catalyze cultural change to further CGIAR OA/OD compliance and public goods mandate. | · Enhanced institutional capacity of partner research organizations<br>· Increased capacity for innovations in partner research organizations<br>· Increased capacity for innovation in partner development organizations and in poor and vulnerable communities<br>· Enhanced individual capacity in partner research organizations through training and exchange | In 2020, Module One continued to develop training materials and guides on achieving and managing FAIR data assets, including licensing, privacy/ethics, and best practices in data management. Center data sprints were encouraged to promote uploading well-annotated datasets to repositories. Module One also supported several workshops and/or training events on ways to improve the FAIRness of CGIAR data assets, and continued to work with Centers and stakeholders, such as the World Bank, to incentivize this. | 2020 - 1.4.7. 2020 - Course materials and webinars for researchers and data managers on best practices for data management and maximizing the FAIRness of CGIAR resources. | Complete | Data assets that are interpretable and interoperable for humans and machines are critical to enhance the impact of research in the agricultural domain and to catalyze innovation and transform agricultural research for development. A course entitled "Best Practices for Open, FAIR, and Ethical Data" was made available via the TechChange platform in 2020 with the goal of strengthening capacity in the CGIAR System and beyond to create, manage, and share research and development data assets that are not only open (i.e., discoverable and downloadable), but also easily interpretable, interoperable, and reusable. Several webinars were also offered on data management-related topics under the auspices of the Information and Data Management CoP of the Module, other CoPs, and the Platform webinar series. | M. Devare, T. Hazekamp. 2020. Best Practices for Open, FAIR, and Ethical Data. Online course accessible from TechChange: https://bigdata-cgiar.course.tc/fair Events section of IDM CoP space on BIG DATA website: https://bigdata.cgiar.org/communities-of-practice/info-data-mgmt/ |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| | | | | 2020 - 1.4.8. 2020 - At least two workshops and/or training events for data and information managers and researchers on ways to render datasets FAIR, including through the use of standards-compliant field books for data collection. | Complete | Two user training events were held on the FAIRification workflow, involving data managers and information specialists from CGIAR and the World Bank. A training on CG Labs was organized for members of the Community for Spatial Information, consisting of CGIAR and external researchers. A demo and hands-on workshop was organized for data managers and biometricians at ICRISAT on the use of AgroFIMS to collect data that is born-FAIR. Several other AgroFIMS discussions were also organized with project teams at CIMMYT and IITA, and documentation and a video developed to help train AgroFIMS users. | AgroFIMS documentation: https://agrofims.org/documentation AgroFIMS video: https://agrofims.org/about |
| **M2** | M2 Outcome: 2.1. CGIAR is more broadly engaged in BIG DATA community. | · Enhanced institutional capacity of partner research organizations<br>· Enhanced individual capacity in partner research organizations through training and exchange<br>· Increased capacity for innovations in partner research organizations | BIG DATA held its annual Convention entirely online, attracting 8,000 registrants and generating massive digital media reach. CoPs grew to more than 5,000 public, private, and non-profit members. 72 external partnerships helped drive digital innovation in research; unlock data sharing across research organizations and public, private, and non-profit partners; and deliver new impact pathways for CGIAR. Of its 2020 partnerships, 30% were from the private sector; 26% universities; | 2020 extended to 2021 - 2.1.5. 2020 - Communities of Practice around geospatial data, socioeconomic data, ontologies, data-driven agronomy, livestock data, and crop modeling establish CoP networks across CGIAR and produce outputs addressing key constraints in data and analytics. | Extended | The CoPs continued to grow and produce rich technical engagement in each of their thematic areas. The CoPs were also a key source of technical depth for the annual Big Data in Agriculture convention. | https://bigdata.cgiar.org/communities-of-practice/ |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M2 | | | national and international research organizations, 14%; and the remainder was a combination of government, non-profits, development banks, and other international organizations. A majority of CGIAR Centers participated in one or more of these alliances. | 2020 - 2.1.6. 2020 - Hold high-level Convention on Big Data in Agriculture, with wide participation by CGIAR and non-CGIAR actors. Establishment of collaborative agreements. | Complete | The CGIAR Convention on Big Data in Agriculture was produced entirely virtually in October 2020. | https://bigdata.cgiar.org/virtual-convention-2020/ |
| | | | | 2020 extended to 2021 - 2.1.7. 2020 - Develop a pre-competitive, pro-commercial multi-stakeholder alliance for food security research, data sharing, and technology matchmaking and transfer. | Extended | A lawyer specializing in Intellectual Property and partnerships advised the Platform on how to structure such an alliance and to develop the formal agreements to constitute it, but this was not yet formalized at the end of 2020. | An internal consulting report with a model MOU is available on request. |
| | M2 Outcome: 2.2. CGIAR increases its capacity to work on priority topics more quickly, more effectively and at greater scale. | · Enhanced institutional capacity of partner research organizations<br>· Enhanced individual capacity in partner research organizations through training and exchange<br>· Increased capacity for innovations in partner research organizations | CoPs grew to more than 5,000 public, private, and non-profit members, and became an important mechanism for CGIAR to remain abreast of digital innovations in their technical research domains. The Platform piloted a "Data Science Academy" with Coursera for a cohort of 20 CGIAR scientists who studied emergent analytic methods (e.g. machine learning) and common software tools for conducting them (R and Python). | 2020 - 2.2.3. 2020 - Identify and produce high priority, high impact new data products and develop a methodology to produce them. Key products will target gender analysis and climate adaptation. | Complete | The CoPs produced several data products that advance digital innovation in agricultural research for development. | https://bigdata.cgiar.org/communities-of-practice/socio-economic-data/<br><br>https://bigdata.cgiar.org/communities-of-practice/ontologies/<br><br>https://bigdata.cgiar.org/communities-of-practice/data-driven-agronomy/ |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M2 | M2 Outcome: 2.3. CGIAR develops as a learning organization. | · Enhanced institutional capacity of partner research organizations<br>· Increased capacity for innovations in partner research organizations | CoPs grew to more than 5,000 public, private, and non-profit members, and became an important mechanism for CGIAR to remain abreast of digital innovations in its technical research domains. The Platform piloted a "Data Science Academy" with Coursera for a cohort of 20 CGIAR scientists who studied emergent analytic methods (e.g. machine learning) and common software tools for conducting them (R and Python). | 2020 extended to 2021 - 2.3.1. Map CGIAR's needs for common big data-related computing and storage infrastructure. | Extended | This effort builds on the digital strategy research conducted by the Platform as well as its overall architecture design work. The Design Working Group for Digital Services has requested that the Platform lead a team on research infrastructure design that will draw from this work and inform and validate a way forward for One CGIAR. | Internal digital strategy and information infrastructure maps available on request. |
| | | | | 2020 extended to 2021 - 2.3.9. 2020 - Design an aligned, pan-CGIAR vision and action plan for using information technology tools, processes, and infrastructure. | Extended | This effort builds on the digital strategy research conducted by the Platform as well as its overall architecture design work. The Design Working Group for Digital Services has requested that the Platform lead a team on research infrastructure design that will draw from this work and inform and validate a way forward for One CGIAR. | Internal strategy and enterprise architecture design documents available on request. |
| | | | | 2020 - 2.3.10. 2020 - Establish shared services that embody an aligned, unified pan-CGIAR information architecture. | Complete | In 2020, the Platform evaluated an array of shared data services it offered for geospatial data, weather data, population weather data and other data. The team decided that the operational costs and value of these external services did not generally justify the subscription costs. To address this, the Platform centralized its shared services offerings into CGLabs, an analytic environment in which anyone with a CGIAR.ORG or IRRI.ORG | Anyone with a CGIAR.ORG email address can create an account in CGLabs, here: https://labs.scio.systems/index.php/user/auth/login |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M2 | | | | | | email address can create an account and assemble data from GARDIAN or other sources, run commonly used analytic software, and share scripts. CGLabs is powered by Globus, a software service actively used by the scientific computing community. | |
| | | | | 2020 - 2.3.11. 2020 - Develop capacity building activities linked to high demand data science skills. | Complete | 43 Convention sessions and webinars from CoPs, a data science bootcamp for CGIAR scientists, and an online course on FAIR data were all offered in 2020. The Geospatial CoP led a pilot of a data science bootcamp on Coursera. The most popular courses among participants were from curated data science topics, such as "Deep Learning with TensorFlow," "AI in Production," and "Statistics for Machine Learning." | https://bigdata.cgiar.org/rss-article/introducing-cgiar-data-science-academy/ |
| | | | | 2020 - 2.3.12. 2020 - Develop and deliver a pan-CGIAR digital strategy for Board approval. | Complete | BIG DATA conducted wide-ranging consultations and internal assessments. Specifically, in collaboration with Accenture Development Partnerships, it conducted more than 160 surveys, 80 semi-structured interviews, 10 internal workshops on digital trends and capabilities, and an extensive literature review. The 160 surveys, conducted in the second half of 2020, were taken from a wide range of people. Specifically, more than 60 CGIAR researchers and other employees from all 15 Centers were included. | Internal document awaiting circulation and validation by the Board. Draft available here: https://cgspace.cgiar.org/handle/10568/113555 |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M2 | | | | | | The remaining survey sample included external stakeholders in non-CGIAR research, development funding and finance, private sector (not-consulting), private sector (consulting), and small-scale and start-up organizations in the AgriTech industry. This research was guided by three overarching questions: What digital trends will transform agriculture in the next 10 years? What should an organization be able to do to navigate these trends? What roles should public interest actors like CGIAR play in the digital agriculture sector? | |
| M3 | M3 Outcome: 3.1 CGIAR shows how data-driven approaches yield results in poverty reduction, enhanced nutrition or environmental benefits. | · Improved forecasting of impacts of climate change and targeted technology development<br>· Optimized consumption of diverse nutrient-rich foods<br>· Increased resilience of agro-ecosystems and communities, especially those including smallholders | | 2020 - 3.1.4. Successful piloting of five Inspire grantees in Africa and South Asia, and up to four new pilot Inspire projects around big data-related innovations assigned. | Extended | Projects from the 2019 award cycle, and scale-up awards from the 2018 cycle, were implemented in 2020. Some of these experienced implementation delays as a result of reduced access to the field during COVID-19 related quarantine, and request no-cost extensions into 2021. Public reporting for all Inspire Challenge projects continues to grow as these projects evolve. | https://bigdata.cgiar.org/inspire/inspire-winners/ |
| | | | | 2020 - 3.1.6. Synthesis of Inspire project successes and failures in 2019; best practice guidance provided. | Complete | The review of 2019 Inspire Challenge data and lessons learned was rolled into a cumulative evaluation product in 2020, under milestone 3.1.10. 2020. | https://cgspace.cgiar.org/handle/10568/113597 |

| Module | Module Outcomes 2022 | Sub-IDOs | Summary narrative on progress against each Module outcome | Milestone | 2020 milestones status | Evidence for completed milestones or explanation for extended, cancelled or changed | Link to evidence |
|---|---|---|---|---|---|---|---|
| M3 | | | | 2020 - 3.1.7. 2020 - 4 start up and 4 scale up grants awarded in 2019. | Extended | Projects from the 2019 award cycle, and scale-up awards from the 2018 cycle, were implemented in 2020. Some of these experienced implementation delays as a result of reduced access to the field during COVID-19 related quarantine, and requested no-cost extensions into 2021. | https://bigdata.cgiar.org/inspire/inspire-winners/ |
| | | | | 2020 - 3.1.8. 2020- Award grants for up to four new start-up Inspire projects around big data-related innovations. | Complete | 7 new startup projects were awarded in 2020, leveraging a combination of W1/W2 and new bilateral funding. | https://bigdata.cgiar.org/inspire/inspire-winners/ |
| | | | | 2020 - 3.1.9. 2020 – Award scale-up fund to up to four successful pilot Inspire projects (from the 2019 grant recipients) around big data-related innovations. | Extended | Scale-up phase projects from the 2018 and 2019 award cycles were implemented in 2020. Some of these experienced implementation delays as a result of reduced access to the field during COVID-19 related quarantine, and requested no-cost extensions into 2021. | https://bigdata.cgiar.org/inspire/inspire-winners/ |
| | | | | 2020 - 3.1.10. 2020 - Synthesis of Inspire project successes and failures, policy documents, best-practice guidance. | Complete | An external evaluation looking at the entire history of the Inspire Challenge was completed, specifically looking at how successful the Challenge was in advancing digital innovation in agriculture, fostering use of CGIAR and other data, development and evaluation of emergent research methods (e.g. Machine Learning) and digital pathways to impact, and in serving as a partnership mechanism that can build the overall innovation ecosystem for digital agriculture. | https://cgspace.cgiar.org/handle/10568/113597 |

## TABLE 6: NUMBER OF PEER-REVIEWED PUBLICATIONS FROM CURRENT REPORTING PERIOD

| | Number | Percent |
|---|---|---|
| Peer-Reviewed publications | 8 | 100.0% |
| Open Access | 7 | 87.5% |
| ISI | 5 | 62.5% |

## TABLE 7: PARTICIPANTS IN CAPDEV ACTIVITIES

| Number of trainees | Female | Male |
|---|---|---|
| In short-term programs facilitated by CRP/PTF | 10 | 58 |
| In long-term programs facilitated by CRP/PTF | 0 | 0 |
| PhDs | 0 | 0 |

## TABLE 8: KEY EXTERNAL PARTNERSHIPS

| Lead Module | Brief description of partnership aims (30 words) | List of key partners in partnership. Do not use acronyms. | Main area of partnership (may choose multiple) |
|---|---|---|---|
| M1 | Develop high-value data products by applying data science and machine learning on data assets available via GARDIAN. | · Alliance of Bioversity and Centro Internacional de Agricultura Tropical regional hub<br>· Centro Internacional de Mejoramiento de Maíz y Trigo<br>· University of California, Davis | · Research<br>· Capacity Development<br>· Policy |
| M1 | Develop the GARDIAN data ecosystem, including CG Labs. | · SCiO - Big Data in Food Systems | · Delivery<br>· Capacity Development |
| M1 | Collaboration to develop pipelines and tooling from GARDIAN to generate model-ready data. | · University of Florida | · Research<br>· Capacity Development<br>· Delivery |
| M1 | Refinement of the "Copo" tool for user-friendly workflow to render data assets FAIR. | · Earlham Institute | · Capacity Development |
| M2 | Development of ontologies and technical data standards | · Yara | · Other |
| M2 | Using supercomputing to train a model for monitoring NO2 emissions from space | · Hewlett Packard Enterprise | · Research |
| M2 | Collaborative model development and in-kind support of computation for large-scale, in-season crop modeling and high-throughput phenotyping | · Google<br>· Digital Green<br>· One Acre Fund<br>· Peat GmbH<br>· MEDIAE Company | · Research<br>· Other<br>· Delivery |
| M2 | Collaborative research on emerging priority topics in digital agriculture and AI | · University of Cambridge | · Policy<br>· Research |
| M3 | Data science competitions as a new open innovation approach for CGIAR using an online competition platform. | · Zindi | · Research<br>· Capacity Development<br>· Other |
| M3 | Co-development and funding of Measuring and Building Resilience Inspire Challenge category | · U.S. Agency for International Development | · Research<br>· Delivery |
| M3 | Co-development and funding of Sustaining Farm Income Inspire Challenge category | · Bill & Melinda Gates Foundation | · Delivery<br>· Research |

## TABLE 9: INTERNAL CROSS-CGIAR COLLABORATIONS

| Brief description of the collaboration | Name(s) of collaborating CRP(s), Platform(s) or Center(s) | Optional: Value added, in a few words |
|---|---|---|
| Collaboration to facilitate the use of the GARDIAN API to pull CGIAR data and publications of relevance into new GENDER Platform web pages. | GENDER | |
| Collaborative work to develop AgroFIMS. | CIP | |
| Collaboration to build awareness of and test BIG DATA tools and services through the IITA-led Excellence in Agronomy initiative. | IITA | |
| Collaboration with UC Davis to develop a data product for assessing the profitability of fertilizer use across Sub-Saharan Africa. | CIMMYT, CIAT | |
| Collaboration to test and refine CG Labs based on high-compute requirements of large climate and vulnerability analysis of the Green Innovation Centres (GIC) target commodity value chains in 16 countries across Sub-Saharan Africa, and India and Vietnam. | CCAFS | |
| Data standards development to unlock gender research enabled by CGIAR data. Collaboration on using call detail records for examine factors related to women's economic empowerment in agriculture with greater frequency and scale than common survey-based methods. Development of human-centered design approaches to digital agriculture that enhance social inclusion. | GENDER | Development of new methods and positioning of human-centered design in the CGIAR portfolio. |
| Co-funding of a project to develop standard operating procedures for using drones for high-throughput phenotyping. Developing programming interfaces with the Enterprise Breeding System to facilitate easier use of crop models and linkages to envirotyping and target product environment analysis. | EiB | More unified pan-CGIAR science infrastructure |
| Serve on the Policy Module of the Genebanks, specifically looking at strategies and approaches for Digital Sequence Information data, development of a potential new data standard linked to the Public Information System under the International Treaty for Plant Genetic Resources for Food and Agriculture. | Genebank | More unified scientific infrastructure, new tools for overcoming governance bottlenecks related to genomic sequence data |
| CIMMTY and CIAT/Bioversity were the CGIAR partners engaged in collaborative model development and in-kind support of computation for large-scale, in-season crop modeling and high-throughput phenotyping linking X, Digital Green, One Acre Fund, Mediae Company, and PEAT. | CIMMYT, CIAT | Digital innovation in research and developing new digital delivery and pathways to impact. |
| Project teams linking researchers with external partner organizations of any type were final awardees under the 2020 Inspire Challenge. Projects from WorldFish, IFPRI, and IITA (second-stage grants awarded in 2019) were also implemented in 2020. | ILRI, ICRAF, WorldFish, IFPRI, ICARDA, CIMMYT, IITA, ICRISAT, BIOVERSITY, CIAT | Development of new digital delivery methods and pathways to impact. |

## TABLE 10: MONITORING, EVALUATION, LEARNING AND IMPACT ASSESSMENT (MELIA)

| Studies/learning exercises planned for this year (from POWB) | Status | Type of study or activity | Description of activity / study | Links to MELIA publications |
|---|---|---|---|---|
| S2957 - Emerging priority topics in digital agriculture | Cancelled | Ex-post adoption study | Research turned out to be more horizon scanning instead of adoption focused. | |
| S3043 - Survey of digital agriculture strategic landscape and CGIAR capabilities | Completed | Synthesis (secondary) study | Part of the digital strategy. The digital strategy research (160 surveys, 80 interviews) led by the Platform generated internal documents that will be released in 2021. Two related public assessments were published in 2020. | https://cgspace.cgiar.org/handle/10568/108402 https://cgspace.cgiar.org/handle/10568/108095 |
| S3884 - CGIAR modeling approaches for resource-constrained scenarios: I. Accelerating crop breeding for a changing climate | Completed | Synthesis (secondary) study | Peer-review publication elaborated in collaboration with different CGIAR centers to present the work CGIAR has done related to the role Crop Modeling has played within CGIAR to accelerate and improve the impacts of breeding efforts. | https://cgspace.cgiar.org/handle/10568/108316 |
| S3885 - CGIAR modeling approaches for resource-constrained scenarios: II. Models for analyzing socioeconomic factors to improve policy recommendations | Completed | Synthesis (secondary) study | Peer-review publication that provides insights into the richness of the socioeconomic modeling endeavors within CGIAR. By linking crop models with economic models and approaches, crop model output can be effectively used as input into socioeconomic modeling efforts for priority setting and policy advice using ex-ante impact assessment of technologies and scenario analysis. The study also highlights the need for interdisciplinary approaches to address the challenges this type of modeling faces. | https://cgspace.cgiar.org/handle/10568/108315 |

## TABLE 11: UPDATE ON ACTIONS TAKEN IN RESPONSE TO RELEVANT EVALUATIONS

| Name of the Evaluation | Recommendation number (from evaluation) | Text of recommendation (can be shortened) | Status of response to this recommendation | Concrete actions taken for this recommendation. | By whom (per action) | When (per action) | Link to evidence |
|---|---|---|---|---|---|---|---|
| N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

## TABLE 13: PLATFORM FINANCIAL REPORT

| | Planned Budget 2020* | | | Actual Expenditure* | | | Difference* | | | Comments |
|---|---|---|---|---|---|---|---|---|---|---|
| | W1/W2 | W3/ Bilateral | Total | W1/W2 | W3/ Bilateral | Total | W1/W2 | W3/ Bilateral | Total | |
| **M1 - Organize** | US$ 1,071,785.00 | US$ .00 | US$ 1,071,785.00 | US$ 994,668.00 | US$ .00 | US$ 994,668.00 | US$ 77,117.00 | US$ .00 | US$ 77,117.00 | The Organize Module,—centered on data standards, management, and analytic infrastructure —completed its year having met its program milestones, within budget. Other programs began to leverage these infrastructures even as the overall budget for them declined, as planned in the BIG DATA proposal. The team sees this as a key sign of progress in helping CGIAR become more data- and digitally-enabled. |
| **M2 - Convene** | US$ 1,628,039.00 | US$ 1,723,839.00 | US$ 3,351,878.00 | US$ 1,827,353.00 | US$ 1,096,863.00 | US$ 2,924,216.00 | US$ -199,314.00 | US$ 626,976.00 | US$ 427,662.00 | The Platform invested its Module Two budget in support of digital adaptation in food systems, highlighting CGIAR's special role in achieving it (specifically showcasing digital investments and dynamism from all CGIAR Centers in the annual Convention) and made a special round of "rapid response grants" to leverage on-the-ground projects for COVID response, recovery, and resilience. |

| | Planned Budget 2020* | | | Actual Expenditure* | | | Difference* | | | Comments |
|---|---|---|---|---|---|---|---|---|---|---|
| | W1/W2 | W3/ Bilateral | Total | W1/W2 | W3/ Bilateral | Total | W1/W2 | W3/ Bilateral | Total | |
| **M3 - Inspire** | US$ 1,625,917.00 | US$ .00 | US$ 1,625,917.00 | US$ 1,230,062.00 | US$ .00 | US$ 1,230,062.00 | US$ 395,855.00 | US$ .00 | US$ 395,855.00 | Several Inspire Challenge projects experienced moderate delays due to lockdown constraints and requested no-cost extensions, resulting in some budget carryover. In addition, some scale-up stage projects given their award in 2019 continued working on their projects in 2020, which was another source of budget carryover. The Module is on track to expend all budget resources by the end of 2021. |
| **Strategic Competitive Research grants** | US$ .00 | US$ .00 | US$ .00 | US$ .00 | US$ .00 | US$ .00 | US$ .00 | US$ .00 | US$ .00 | No strategic competitive research grants were awarded to the Platform in 2020. |
| **Platform Management & Support Costs** | US$ 412,336.00 | US$ .00 | US$ 412,336.00 | US$ 327,868.00 | US$ .00 | US$ 327,868.00 | US$ 84,468.00 | US$ .00 | US$ 84,468.00 | The Platform Secretariat and internal support functions at the Alliance of Bioversity and CIAT reached their programmatic goals within budget in 2020. |
| **Platform Total** | US$ 4,738,077.00 | US$ 1,723,839.00 | US$ 6,461,916.00 | US$ 4,379,951.00 | US$ 1,096,863.00 | US$ 5,476,814.00 | US$ 358,126.00 | US$ 626,976.00 | US$ 985,102.00 | |

# TABLE ANNEXES

| FP | Detailed Annex |
|---|---|
| **M1: Organize** | The GARDIAN data ecosystem offers technical tools designed to help researchers and data managers make CGIAR's data assets open and FAIR at every stage in the data management lifecycle. To facilitate participation in this one-stop knowledge base and answer "How do I do this?", GARDIAN made a data management toolkit available in 2020 with easy-to-use resources that allow users to more easily make their data assets open, FAIR, and responsibly managed. The toolkit includes links to data standards (e.g., the CG Core Metadata Schema, ontologies) as well as guides on responsible and FAIR data management plus workflows to easily make research outputs FAIRer. These tools help researchers and data managers more easily contribute to an increasingly open, interoperable and reusable data pool that can be leveraged to quickly and responsively develop insights that can help transform agriculture. Additionally, the Agronomy Field Information Management System (AgroFIMS) enables data digitally collected in the field to be tied to data standards at the time of collection, making it "born-FAIR" data. Development efforts on AgroFIMS in 2020 included a streamlined interface and other options. Its developers will continue to respond to user needs and demand for this tool through the Excellence in Agronomy Initiative (EiA).<br><br>The heart of the ecosystem is the GARDIAN discovery portal, which enables all CGIAR data and publication repositories to be searched along with those managed by a growing group of strategic partners including public agricultural research agencies, development funders, and the World Bank. In 2020, GARDIAN was used as a source of CGIAR data and/or publications by the FAO's Agris which provides access to publications for a wide range of partners; and by the GENDER Platform and the Alliance of Bioversity and CIAT (Alliance) as an enhancement to their webpages. Each new partner that joined the network to share data—be it an external strategic partner or a CGIAR program—further reinforces CGIAR's role as a trusted, capable broker or intermediary of open and FAIR agricultural data.<br><br>The Collaborative GARDIAN Labs (CG Labs) analytical environment is critical to this goal. In 2020, it increased the number of large datasets, computational power, and crop simulation model pipelines offered to users. These enhancements enabled more "plug-and-play" functions similar to those biomedical researchers have enjoyed for some decades and that have helped transform medical therapies. CG Labs allows researchers to not only find data via GARDIAN, but also to save it through CG Labs, upload data from other sources, and couple that with large climate or weather data that is part of the CG Labs bundled services. Thanks to a collaboration with the University of Florida's Agricultural Model Intercomparison and Improvement Project (AgMIP), data can be input to crop models commonly used by CGIAR and other agricultural researchers (e.g., DSSAT, WOFOST, EcoCrop), then used in spatial analysis and the development of maps using any one of hundreds of available pre-packaged R packages. CG Labs also allows built-in interaction with programming scripts that can be easily shared and made available for all to use. It was used in 2020 by members of CGIAR's Community for Spatial Information, Climate Change, Agriculture and Food Security (CCAFS), the Excellence in Agronomy Initiative (EiA) and other CGIAR and World Bank researchers for high compute, data-intensive tasks.<br><br>The Organize Module also leveraged semantic web technology to enhance the CGIAR Expert Finder (alpha version), which showcases CGIAR's research expertise and global reach. The Expert Finder allows CGIAR leadership and researchers, funders, and other stakeholders to quickly find who works where across the System based on keyword searches (e.g., food security, mechanization). It also allows researchers to easily find new collaborators on a given topic and/or in particular geographic areas or Centers, and it enables Center leads to easily get an interactive view of entities working around a particular theme. Expert Finder can also provide an interactive mapping of CGIAR capabilities around any specific research theme or across themes. The tool currently compiles publicly available data (primarily publications and datasets) on CGIAR researchers, but it could be greatly improved were individual researchers encouraged to edit any non-authoritative information in the system, such as research summaries or geographic areas of expertise. It was used in 2020 as a source of information on people involved with the GENDER Platform, and, as an easily sliced resource, is being considered by a number of Centers as a potential content source for their web pages. |

| FP | Detailed Annex |
|---|---|
| **M2: Convene** | BIG DATA was presented with a unique opportunity in 2020 to "walk the talk" on agile, adaptive, digitally-enabled collective action. For the Platform's fourth annual convention, it transitioned its annual convention to be an inclusive, accessible, and fully online event. The event theme, Digital Dynamism for Adaptive Food Systems, examined food system resilience and highlighted how digital tools and technologies can help various food system actors sense, respond, and (re)build better systems in global food security crises.<br><br>The 2020 Convention and recorded the greatest amount of CGIAR-wide engagement to date. Every CGIAR Center submitted an application to the Inspire Challenge and almost every Center was represented in the 10 finalist projects. With the Platform's support, CGIAR Centers submitted content that included 11 short videos offering a glimpse of digital dynamism at a Center level, as well 15 tech demos highlighting some of these centers' most cutting-edge digital innovations.<br><br>This was the most inclusive Convention to date. It provided an opportunity to showcase how each of the global Centers are employing digitally-enabled, dynamic methods to combat global food security challenges originating in current crises. The free virtual event attracted more than 1,300 active participants from around the world (75% of them external to CGIAR) who engaged in sessions and chats as well as on discussion boards around topics related to digital adaptation in global food security. The virtual format also enabled the Platform to engage a wider array of speakers. Our 150 speakers included both high-profile industry experts as well as digitally-enabled farmers from three regions. Half of these speakers were women.<br><br>The 2020 Youth In Data Workshop, which was conducted in parallel with the Convention, enrolled 130 youth from 49 countries. These young digital innovators interacted with and interviewed Convention experts and participants, resulting in them writing five blogs that were published on the BIG DATA website. They were key in the Platform's success in reaching more than 14 million people via the convention #BDPGLOBAL2020 tag across Twitter alone. |
| **M3: Inspire** | Some important results from Inspire Challenge projects achieved in 2020 include:<br><br>PlantVillage Nuru played a key role in responding to the 2020 locust swarms in East Africa by adapting its groundbreaking, AI-informed approach to pest and disease monitoring. Borrowing from the Plantvillage Nuru application blueprint, the team developed a new smartphone application in just one month for monitoring locusts across the region. The 240,000+ locust records captured via the eLocust3m app enabled countries to systematically act to control the swarms. Since February 2020, the FAO estimates that this effort in East Africa has averted the loss of commercial agricultural products valued at US$1.5 billion—saving the livelihoods of 34 million people.<br><br>Hungry Cities leveraged its high frequency data on the metropolitan Nairobi food system to get fresh fruit and vegetables (FFV) to the people who needed them the most during the COVID-19 pandemic. In June 2020, the project's data showed that 90% of households in the Nairobi slums were experiencing dire food insecurity, including reduced FFV consumption, primarily due to reduced household incomes. The project's data-driven insights led Twiga Foods, a leading agricultural produce supplier in Kenya, to reduce its FFV prices in its markets that serve hundreds of thousands of low-income consumers.<br><br>Another project has become an instrumental partner to the General Statistics Office of Vietnam in assessing COVID-19-related food security shocks at a national level. Since 2019, the team has continually monitored informal food flows in five traditional markets in Hanoi, creating a valuable dataset of more than 200 million total data points related to the impacts of and actor behavior changes before, during, and after the first and second waves of COVID-19. The early analysis reveals a sharp decrease in the average number of people seen daily in the urban markets in March 2020, compared to 2019 baseline numbers, and 38% of previously frequent visitors—those who had typically visited at least five times a week pre-COVID—came less often or totally avoided the markets months after the pandemic's onset. |

| FP | Detailed Annex |
|---|---|
| **M3: Inspire** | Relevance to Covid-19 (300 words):<br>The COVID-19 crisis presented BIG DATA with the opportunity to demonstrate that digital tools, methods, and partnerships provide the critical agility and adaptability needed to manage food security shocks. As a result, the Platform realigned some internal budgets to run the Rapid Response grant challenge in which previous awardees could seek additional funding to leverage their existing projects and partnerships to implement response, recovery, and resilience interventions for mitigating local food security impacts.<br><br>The awards and their projected impacts are as follows:<br><br>Rapid diagnostics of COVID-19 farming impacts developed a mobile phone-based panel survey approach to assess COVID-19 impacts on farming, which is being tested at scale across Kenya through the "Let it Rain" campaign. In 2020, the team completed two surveys with more than 2,000 farmers on COVID-19 impacts. Two additional surveys will be conducted in 2021. The complete set of findings will inform targeted, digital advisories for farmers and programmatic responses through info briefs.<br>Herd opportunity will deploy an interactive digital course to train livestock farmers on how to prevent the spread of infectious diseases, including COVID-19, on and off the farm. The team will pilot the course in Kenya and aims to reach between 5,000 and 10,000 livestock farmers within the first four months. Within two years, the team aims to scale up its training to more than 250,000 farmers across sub-Saharan Africa. Using data from the interactive course, the project will also generate live data on hygiene practices and infection risk among livestock farmers through an open access data dashboard.<br>Eyes on the ground for agricultural microcredit aims to unlock agricultural microcredit to farmers in Odisha, India, by enabling a microfinance lender to monitor the loans its makes using satellite and smartphone imagery. The team aims to pilot this approach by issuing US$400 loans to 450 farmers in 50 randomly-selected villages. |

CGIAR

# Platform for Big Data in Agriculture